

## **A FCSH na Web: construindo a memória da presença online**

**Ricardo Basílio**

### **1. Definição do projecto**

1.1 O problema do desconhecimento - o que é e para que serve um arquivo da Web

As instituições académicas dispõem de uma grande variedade de recursos para integrar nos seus processos investigação: livros e revistas na versão impressa, livros e artigos em versão electrónica, bases de dados e também o Arquivo.pt que tem divulgado o seu serviço. Os serviços da biblioteca procuram dar-lhes visibilidade, rentabilizando assim o investimento que implicam.

Todavia, o material preservado pelos arquivos da Web, que consiste fundamentalmente em versões datadas de páginas da Web, ainda é uma novidade. Não é de admirar haver investigadores que não conhecem o Arquivo.pt, e outros que, embora conhecendo, não o integram na sua investigação.

Se a utilização dos documentos em formato electrónico é preterida em favor dos documentos impressos, muito mais isso acontece quando se trata de sítios Web do passado. O que fazer com eles, qual a sua utilidade para uma investigação e em que áreas são mais pertinentes?

Para levar cada vez mais investigadores a procurarem a resposta para estas questões, e antes de mais, para dar a conhecer o que é e como funciona a interface do Arquivo.pt, pensou-se numa demonstração simples e acessível, aplicada ao tema da memória institucional que está contida nos sítios Web do passado.

Assim surgiu uma proposta, na perspectiva do curador digital que quer dar visibilidade a um novo recurso ao dispor de uma comunidade de utilizadores.

1.2 Uma exposição sobre memória de uma instituição

“Memória da FCSH na Web: presença *online* da Faculdade de Ciências Sociais e Humanas 1996-2016” é um caso de uso do Arquivo.pt, limitado ao contexto institucional da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa (FCSH-UNL).

O conteúdo da exposição são os sítios Web que foram preservados pelo Arquivo.pt e que têm relação com a FCSH-UNL e com os centros de investigação. Estes são de três tipos: *sítios Web* institucionais, *sites* de projectos e *sites* de eventos (colóquios, seminários, etc.), que constituem parte da memória *online* da FCSH-UNL.

O foco da atenção e o material principal do projecto são os sítios Web do passado, mais concretamente, as versões do passado preservadas pelo Arquivo.pt.

## 1.2 Uma oportunidade num ano de celebrações

A FCSH criou o primeiro sítio Web há 20 anos e tem um mais de uma centena de sítios Web preservados pelo Arquivo.pt. Em Novembro de 2017, celebra 40 anos da sua fundação.

Actualmente, a FCSH tem 12 departamentos e 16 centros de investigação, também referidos como unidades I&D. Algumas unidades tornaram-se inter-universitárias. Chegou a ter mais de 30 unidades que depois se fundiram em outras maiores. Algumas chegam a enquadrar cerca de 300 investigadores (por exemplo, Instituto de História Contemporânea - IHC).

As áreas de ensino e de investigação Faculdade coloca-a no âmbito das humanidades digitais, uma área que tem merecido a atenção de organismos responsáveis pela política europeia para a Ciência e que está em desenvolvimento.

No Reino Unido, por exemplo, emergências das humanidades digitais esteve na origem do projecto BUDDAH (Big UK Domain Data for the Arts and Humanities), o qual consistia em pedir a investigadores para explorarem as possibilidades do arquivo da Web, apresentando depois os resultados dessa experiência. No projecto estiveram envolvidas várias instituições o que demonstra a natureza forçosamente colaborativa de projectos semelhantes: uma biblioteca, universidades, centros de investigação e uma associação de Humanidades Digitais (<https://buddah.projects.history.ac.uk/>).

Este projecto inspira-se no projecto BUDDAH e foi com ele que o Arquivo.pt quis marcar 10 anos de actividade (2007-2017). Assim, lançou uma chamada para projectos ou ideias que fizessem uso do serviço e demonstrassem a sua importância para a investigação em diversas áreas, nomeadamente nas Humanidades.

No caso deste projecto de exposição virtual, as questões de investigação foram as seguintes:

Como fazer uma exposição representativa da presença da FCSH na Web ao longo dos anos, a partir da informação preservada pelo Arquivo.pt desde 1996?

Como criar algo que permita a muitos investigadores da Faculdade terem um primeiro contacto com o Arquivo.pt e, para aqueles que já o tiveram, verem um exemplo do que é possível fazer utilizando este serviço?

Foi na perspectiva de bibliotecário que o primeiro esboço para uma exposição *online* foi imaginado. Um bibliotecário pretende preparar recursos para os investigadores e mostrar, o melhor possível, que estes existem.

### 1.3 Destinatários do projecto

O projecto destina-se, imediatamente, à comunidade académica da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa e, em geral, a todo o utilizador com interesse em conhecer os sítios Web do passado desta instituição e a sua história.

Foram especificados alguns tipos de potenciais visitantes da exposição *online*. Para isso foram criadas *personas*, ou seja, perfis de utilizador, que ajudaram a definir os conteúdos da exposição e a testar o funcionamento da exposição do ponto de vista da usabilidade.

Assim, este projecto destina-se a utilizadores com os seguintes perfis:

- docentes e investigadores das Unidades de Investigação da FCSH;
- bibliotecários e gestores de ciência e tecnologia da Faculdade;
- alunos de todos os graus (licenciatura, mestrado e doutoramento)
- pessoas com alguma relação com o FCSH (antigos alunos, familiares)

Os quatro perfis de utilizador têm em comum uma forte relação com a Faculdade e uma curiosidade natural acerca de referências passadas. A exposição deve por isso manter-se num nível de informação genérico quanto aos conteúdos e simples em relação aos elementos gráficos.

### 1.4 Definição dos objectivos do projecto

- O primeiro é dar visibilidade à memória da FCSH na Web que se encontra nos sítios Web que o Arquivo.pt tem preservado e que, desde 1997 foram produzidos pela Faculdade, pelos seus centros de investigação e outras entidades da comunidade académica (departamentos, grupos de alunos).
- O segundo decorre do tipo de projecto. Este é um caso de uso e como tal tem um carácter exemplar. Espera-se demonstrar o que se pode fazer para construir e perpetuar a memória de uma instituição utilizando como matéria prima os sítios Web preservados por um arquivo da Web. Há, portanto, um objectivo implícito de chamar a atenção dos investigadores para as potencialidades do Arquivo.pt. Se a preservação de sítios Web é útil para documentar aspectos da presença da Faculdade na Web ao longo dos anos, em quantas outras coisas poderá ser útil? Cabe aos investigadores das diversas áreas científicas, especialmente das Humanidades, descobrir por si próprios para que pode servir o enorme volume de informação contido num arquivo da Web. Diversas iniciativas internacionais sobre preservação digital insistem na importância dos materiais nascidos digitais, tais como os sítios Web, para o conhecimento da sociedade dos últimos anos.

- O terceiro objectivo é dar um contributo para a preservação dos sítios Web da Faculdade. Ao mostrar a utilidade e o potencial interesse dos sítios Web do passado também se motiva para preservar a informação *online* que está a ser produzida actualmente.

## **2. Preparação da exposição *online***

### 2.1 Compreendendo o que é um arquivo da Web

Um dos primeiros empenhos do projecto foi a aquisição de conhecimentos mais específicos sobre o Arquivo.pt, ou seja, a formação. Esta serviu para compreender o processo de recolha de sítios Web e as diversas tecnologias envolvidas. Tratou-se de uma formação de nível médio, e pontualmente de nível avançado sempre que surgiram questões relacionadas com sítios Web da exposição.

Sabe-se, portanto, que no Arquivo.pt a recolha é feita por um *crawler* que passa pelos servidores, copiando os ficheiros que formam um sítio Web, bem como as ligações entre eles. O *crawler* faz uma cópia do que pode, ou seja, com limitações, umas impostas pelo administrador do *site* através da configuração do ficheiro *robots.txt*, outras pela política de recolha do Arquivo.pt.

Como resultado, o material disponível para reconstituir os sítios Web é inevitavelmente fragmentado. Assim, por exemplo, se o *crawler* não tiver acesso aos ficheiros CSS, responsáveis pelo aspecto (*design*) ou se lhe for vedado o acesso às imagens, a reconstituição do sítio Web fica incompleta. Existem também tecnologias usadas no passado para a publicação de conteúdos Web, as quais pelas suas características de baixa preservabilidade nem sempre conseguem ser reproduzidas pelo Arquivo.pt.

Exemplo disso são os conteúdos em Flash, um *software* proprietário que teve várias versões ao longo do tempo. Para a leitura desses ficheiros nos *browsers* foi criado o *software* correspondente com as sucessivas actualizações. O problema surge quando a versão do *software* que produziu um ficheiro não é compatível com a versão que está no *browser* do utilizador. Nesse caso, a apresentação do conteúdo fica comprometida. Assim, encontramos sítios Web em Flash que, embora guardados pelo Arquivo.pt, não funcionam em dispositivos actuais, por lhes faltar a versão certa para a sua leitura. Para testar este caso, experimente-se o sítio Web do Instituto de História da Arte da FCSH, de 2004 (IHA).

O Arquivo.pt apresenta as versões organizadas cronologicamente, permitindo comparar páginas em datas diferentes. Deve ter-se em conta que nesse nível de apresentação ocorrem, por vezes, saltos no tempo, ou seja, passagens inesperadas de uma data para outra, que exigem mais atenção.

### 2.2 O que se entende por presença *online*?

### 2.2.1 A Web é mais do que tecnologia

Esta questão preliminar não foi fácil de responder, porque os sítios Web estão associados a vários aspectos em simultâneo: há uma tecnologia que lhes dá um determinado aspecto num monitor de computador ou de um dispositivo móvel, a um nome e endereço que por vezes mudam e, finalmente, a uma entidade responsável pelo seu conteúdo que também muda ao longo do tempo.

Do ponto das tecnologias da informação, um sítio Web é um conjunto de ficheiros de diversos formatos, escritos em diversas linguagens, colocados num servidor Web. Estes ficheiros estão de tal modo ligados entre si e suportados por outras tantas tecnologias que chegam a produzir um conteúdo, legível, acessível através da Web, num *browser*.

Se tivermos em conta apenas o aspecto tecnológico, e recorrendo ao exemplo do sítio Web do Instituto de História da Arte (IHA) no Arquivo.pt, dir-se-ia que houve um *site* do IHA entre 2008 e 2012, outro entre 2012 e 2014 e ainda outro, de 2014 em diante feito em Wordpress, porque houve mudanças no *design*, no nome e no sistema utilizado.

A tecnologia e a aparência mudaram, mas continuou a ser fundamentalmente o sítio Web do Instituto de História da Arte. Não foram vários sítios Web, mas apenas um, criado para a comunicação institucional do IHA.

Significa que uma perspectiva meramente tecnológica sobre os sítios Web é insuficiente. Assim sendo, decidiu-se adoptar a perspectiva organizacional para definição de sítio Web neste projecto. Um sítio Web define-se pela entidade que o criou.

Um sítio Web é um canal de comunicação criado por uma entidade (pessoa, grupo ou organização) para um determinado fim, que pode ser tão amplo como a comunicação institucional ou tão específico como a cobertura de um evento. Tem por referência uma entidade prevalente ao longo do tempo que assume a responsabilidade principal pela sua criação, manutenção e desactivação. Inclui responsabilidades específicas, que podem ser partilhadas, tais como a edição e as condições de acesso ou de reutilização dos conteúdos. Essa entidade é fio condutor que cria a interligação entre as diversas versões de um sítio Web ao longo do tempo. Assim, por exemplo,

Assim, por exemplo, pode considerar-se um sítio Web a página dedicada ao IHA em 2002, no *site* da Faculdade? É difícil responder afirmativamente. (Cf. <http://arquivo.pt/wayback/20020614233300/http://www.fcsh.unl.pt/Iha/default.htm>)

Convém perguntar antes qual era a função dessa página do ponto de vista da comunicação do IHA. Trata-se de uma página que se manteve inalterada em todas as versões disponíveis no Arquivo.pt, ou seja, entre 2000 e 2005. Dá a impressão

que foi criado um espaço no sítio Web da Faculdade – uma página para cada unidade de investigação - , que esse espaço foi preenchido com alguma informação, mas sem ligação imediata com o desenvolvimento da unidade de investigação em causa. Neste sentido, pode considerar-se esta página Web isso mesmo, uma página, não um sítio Web. Pode acrescentar-se ainda que tal página serviu para apresentar sumariamente o IHA. O sítio Web foi criado mais tarde, em 2008, e esse é o que vai ser incluído na exposição *online*, não por ter muitas páginas, mas por ser um canal comunicação activo (<http://arquivo.pt/wayback/20080219035604/http://iha.fcsh.unl.pt/> ).

Não existe uma definição determinística. O exemplo acima não resolve todas as questões de outras páginas que parecem *sites* ou de *sites* que parecem páginas (incluídas dentro do *site* da FCSH, por exemplo). Por isso a análise foi feita caso a caso, em função do objectivo da exposição de apresentar a memória *online* das diversas entidades de uma organização.

### 2.2.3 Os elementos da presença *online* de uma organização ao longo do tempo

Se no tópico anterior se centrou a atenção na organização criadora de um determinado sítio Web, neste refere-se vários elementos – IP, *hostname*, título, conteúdo e código-fonte - que foram utilizados no processo de pesquisa e análise dos sites do passado da FCSH, a fim de encontrar as suas versões preservadas no Arquivo.pt :

- 1) O IP é uma sequência numérica e hierárquica que indica o acesso a um servidor Web. Por vezes, essa é a única forma disponível para encontrar *sites*, por exemplo, quando deixou de haver um *hostname* associado. Durante a pesquisa encontrou-se pelo menos um sítio Web de um projecto nessas condições.
- 2) O *hostname* é um nome que nos dá acesso a um servidor Web, em vez de uma sequência de números e por isso é mais fácil de memorizar. Na pesquisa no Arquivo.pt utilizou-se o *hostname* como expressão de pesquisa, pois as organizações tendem a usar o *hostname* para representar relações de pertença ou enquadramento institucional. Por exemplo, *iha.fcsh.unl.pt* expressa uma relação de pertença e um enquadramento hierárquico entre várias entidades: IHA, FCSH e UNL. Esta prática não é regra, pois nem sempre o *hostname* acompanha a dinâmica das organizações. Estas mudam e o *hostname* continua o mesmo, por questões práticas. Tal é o caso do CHAM que, ao longo da sua história, tem mantido o mesmo *hostname*. Outras vezes as unidades de investigação dispensam usar um *hostname* com enquadramento hierárquico, e optam por soluções mais flexíveis. Assim, por exemplo, o IHA mudou o seu *hostname* de *iha.fcsh.unl.pt* para *institutodehistoriadaarte.wordpress.com*, a partir de 2014.
- 3) O título do *site* é um importante identificador. Não só se encontra no topo da página, em caracteres maiores, como também habitualmente informa sobre o enquadramento institucional. Assim, por exemplo, o sítio Web do IHA, apesar

de utilizar o *hostname* [institutodehistoriadaarte.wordpress.com](http://institutodehistoriadaarte.wordpress.com), o qual nada diria sobre a sua relação institucional com a Faculdade, apresenta aos seus utilizadores um título que diz tudo: "Instituto de História da Arte – Faculdade de Ciências Sociais e Humanas".

- 4) O conteúdo das páginas é a fonte de informação que pode esclarecer sobre a relação de um sítio Web com uma entidade, quando nem o *hostname* nem o título são suficientes. Por exemplo, quando surgem dúvidas sobre a relação de uma Unidade I&D com a Faculdade, o que acontece algumas vezes quando estas alteram a sua configuração (fundindo-se, integrando-se em redes inter-universitárias) é necessário a consulta dos estatutos ou de outros textos auto-descritivos presentes no site. Se, ainda assim, a interpretação não for clara, convém consultar os responsáveis pela entidade que se quer representar na exposição virtual.
- 5) Finalmente, o código-fonte das páginas pode ser útil para ter alguma informação sobre um sítio Web, sobretudo quando faltarem outros elementos. Veja-se por exemplo as tags do cabeçalho HTML ("Title") e outras com meta-informação.

Os elementos apresentados, por si só, não constroem a memória da presença online de uma organização. Daí a importância do profissional de informação, cuja função que é relacionar os dados encontrados, analisar as versões dos sítios Web ao longo do tempo, compreender o enquadramento institucional e apresentar os resultados de forma compreensível para quem quer explorar a memória institucional. Todo este processo não é mais do que curadoria digital.

## 2.3 Selecção dos *sites* para descrição da memória online da FCSH

### 2.3.1 Questões de representação institucional dos *sites*

Em princípio deveria ser fácil dizer quais são os *sites* de uma instituição. No caso da FCSH as unidades I&D têm sofrido alterações na sua estrutura e na sua identidade ao longo dos anos e nem sempre essas mudanças estão claramente reflectidas nos respectivos sítios Web.

Há as que optam pelo encerramento do *site* anterior, mudam o nome e o domínio e criam um *site* completamente novo para representar uma nova fase.

Outras, porém, preferem manter o nome e o domínio antigos do seu *site*, mesmo quando passaram a integrar outras organizações que antes eram distintas e mesmo quando mudam de nome. Tal aconteceu, por exemplo, no caso do CHAM, que tem mantido o *hostname* [cham.fcsh.unl.pt](http://cham.fcsh.unl.pt) ao longo dos anos. Inicialmente, chamava-se Centro de Estudos de Além-Mar. Em 2013, passou a chamar-se Centro de Estudos d'Aquém e d'Além Mar, ano em que passou a integrar investigadores de outros centros (Centro de História da Cultura, Centro de Estudos Históricos e Instituto

Oriental). Em 2017, o CHAM apresenta-se com o nome simplificado – CHAM Centro de Humanidades.

Estas questões são pertinentes, quando se pretende identificar um sítio Web para o descrever. É necessário apurar se a identidade de uma organização mudou tanto que passou a ter outro âmbito.

O *crawler* é cego em relação a estas variações e baseia as suas recolhas no *hostname*. As mudanças de estrutura ou de identidade de uma organização são-lhe indiferentes. A única coisa que se pode saber olhando para as versões recolhidas no Arquivo.pt é que uso foi dado a determinado *hostname*.

### 2.3.2 Critérios para incluir um site na memória institucional

A inclusão de sites na exposição seguiu os seguintes critérios:

- se a entidade responsável for a Faculdade ou um dos seus centros de investigação;
- se é um site de um projecto em que a Faculdade ou as suas unidades I&D participam como entidades promotoras ou coordenadoras, mesmo que se trate de um projecto em parceria com outras instituições;
- se é um site de um evento promovido pela Faculdade ou por uma das suas entidades (unidades de investigação, departamentos, grupos de alunos, etc), mesmo que esse evento seja organizado em parceria com outras entidades;
- se o *hostname* expressa uma relação com a Faculdade ou uma das suas entidades (ex. iha.fcsh.unl.pt);
- se o título de um sítio Web indicar a pertença à Faculdade ou a relação com das suas entidades (ex., institutodehistoriadaarte.wordpress.com);
- se o conteúdo do site indicar a relação com a Faculdade ou uma das suas entidades sobretudo nos sites de projectos e de eventos realizados em parceria com entidades externas;

Por princípio não foram excluídos sites de projectos e eventos inter-institucionais, apenas por não serem exclusivos da Faculdade ou dos seus centros. Estes são uma parte importante da dinâmica institucional e tem sentido que sejam considerados no processo de recolha.

Em todo o caso, convém verificar junto dos responsáveis por cada *site* se há obstáculos à sua inclusão como fonte da memória institucional. Na descrição, deve ser explicada a responsabilidade e a participação da FCSH ou das suas diversas entidades.

- se o *site* se encontra preservado no Arquivo.pt



Na exposição virtual, apresenta-se os *sites* que se encontram preservados no Arquivo.pt. Os que não estão no Arquivo.pt são excluídos. Desta forma pretende-se reforçar de que se trata de um projecto dedicado à memória.

Não se trata apenas de mostrar *sites* do passado. *Sites* do passado também os há ainda acessíveis através da Web, mas sem a possibilidade de analisar a sua evolução ao longo do tempo.

A exposição *online* quer mostrar a especificidade do *Web archiving* para a preservação digital. Preservar um *site* num sistema como o Arquivo.pt não é a mesma coisa do que armazenar localmente uma cópia do *site* ou manter o *site online*. Se os utilizadores puderem compreender essa diferença, a exposição cumpre um dos seus objectivos.

#### 2.4 Folha de recolha dos dados para a exposição

Uma vez definidos os critérios de selecção de sites para a exposição foi criada uma folha para registar diversas informações durante o processo de pesquisa e análise das versões preservadas no Arquivo.pt. Nesta fase foi recolhida mais informação do que a que é apresentada na exposição *online* (Tabela 1).

Tabela 1: Dados a recolher na pesquisa e análise dos sites no Arquivo.pt

DADOS PARA DESCRIÇÃO DO SÍTIO WEB PRESERVADO	O que se pretende	Elementos utilizados
Nome	Nome ou título do sítio Web para apresentar na exposição	Sim
Contexto institucional	A quem pertence ou que entidade está nele representada. No caso dos <i>sites</i> de projectos e eventos um <i>site</i> pode pertencer a mais do que uma entidade.	Sim
<i>Hostname</i>	<i>hostname(s)</i> , que um sítio Web de uma entidade teve ao longo do tempo e indicação do intervalo temporal correspondente	Sim
Versões guardadas no Arquivo.pt	<i>Links</i> para as versões guardadas pelo Arquivo.pt, um por cada <i>hostname</i>	Sim
Imagem de destaque	<i>Printscreen</i> de uma página preservada para utilizar como imagem na exposição, legenda e URL da página	Sim

História do sítio Web	Nota descritiva das mudanças mais notórias ao longo do tempo, associando sempre o URL para referência.	Sim
Momentos para recordar	Seleção de momentos que se crê "interessantes" para a história da Faculdade ou do centro de investigação.	Sim
Questões	Registo de questões por resolver sobre a história do <i>site</i> (pertença a uma entidade, relação com projectos eventos, etc.).	Não
DADOS PARA A CRIAÇÃO DE RELAÇÕES ENTRE ITEMS	<i>Descrição do que se pretende</i>	
Relacionado com	Levantamento de <i>sites</i> de projectos e de eventos relacionados, incluindo também os que ainda não estão a ser preservados. Os <i>sites</i> que estão no Arquivo.pt são referidos com o URL completo dessa interface.	Não
Tipo de sítio Web	3 tipos: <i>site</i> institucional, <i>site</i> de um projecto, <i>site</i> de um evento (para posterior classificação)	Não
Áreas temáticas	Antropologia, Arte, Comunicação, Filosofia, Geografia, História, Línguas, Literatura, Música, Sociologia (para posterior classificação)	Sim
Contém	Anotação de conteúdos encontrados nas versões preservadas com potencial interesse para os investigadores: artigos em texto integral, resumos, relatórios, notícias, programas de investigação.	Não
ESTADO DO SÍTIO WEB ACTUAL	<i>Descrição do que se pretende</i>	

Versão actual ( <i>live Web</i> )	URL para a versão actual de um sítio Web	Não
Grau de arquivabilidade ou condições de preservação	Verificação das condições de preservação do sítio Web actual – utilizando a ferramenta <a href="http://archiveready.com">archiveready.com</a>  Indicar o URL resultante da verificação.	Não
Questões	Recomendações úteis para a Faculdade e os centros de investigação no sentido de melhorarem a preservação dos seus sítios Web.	Não

A informação recolhida é útil, embora nem toda se utilize imediatamente na exposição virtual, por diversas razões. Introduzir pontos de acesso que não sejam para páginas do passado pode criar confusão no utilizador. Deve ficar claro que se trata de *sites* preservados, que estes são reconstituídos e reproduzidos por um sistema de preservação digital. Por essa razão a referência aos sítios Web actuais foi excluída.

Pela mesma razão, foram excluídos os sítios Web relacionados que não se encontram preservados. O que não está no Arquivo.pt não entra na exposição sobre a memória da FCSH. Fez-se uma lista desses *sites*, normalmente de domínios .com, .net e outros, e enviou-se ao Arquivo.pt com a sugestão de que devem vir a ser preservados, para que não se perca a teia de relações entre os centros de investigação e os *sites* de projectos e de eventos.

Os dados para a criação de relações entre itens, “tipo de sítio Web” e “áreas temáticas”, têm uma aplicação diminuta na exposição, pois o número de itens expostos é pequeno. Porém podem vir a ter utilidade, quando o número de itens for maior e houver vantagem em criar pequenas agregações usando categorias e *tags* no Wordpress.

## 2.5 Metodologia utilizada para pesquisar no Arquivo.pt

Utilizou-se o *hostname*, como expressão de pesquisa na interface do Arquivo.pt. No caso dos sítios Web que mudaram de nome ou de URL ao longo da sua história, também esses dados foram tidos em conta e para cada um foi feita uma pesquisa (ex. [iha.fchs.unl.pt](http://iha.fchs.unl.pt) e [institutodehistoriadaarte.wordpress.com](http://institutodehistoriadaarte.wordpress.com)).

Esta forma de pesquisa conduz-nos a uma página com tabela de versões ao longo do tempo (Fig. 1). Esta é a referência principal para analisar um sítio Web ao longo do

tempo. Voltando metodicamente à tabela mantém-se a sequência cronológica na exploração das versões e toma-se a página de entrada como referência.



ARQUIVO.PT

iha.fcsh.unl.pt

entre: 01/01/1996 e: 31/12/2016

Ver resultados que contém o texto: 'iha.fcsh.unl.pt'

### Tabela de versões

21 versões de iha.fcsh.unl.pt

1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
												19 Fev	21 Mai	31 Mai		23 Jan	5 Nov	26 Set	16 Abr	22 Mai
												11 Mar	23 Jun	6 Jun		23 Jan	7 Nov	30 Set	23 Abr	
												21 Out	24 Set							
												22 Out	25 Set							
													17 Dez							
													18 Dez							

Fig. 1: Pesquisa na interface do Arquivo.pt pelo *hostname* e tabela das versões.

Na tabela das versões começou-se pela data mais antiga da tabela cronológica, que nos indica quando o sítio Web começou a ser preservado (Fig 2). Analisou-se o sítio Web a partir da *homepage* e qualquer incursão por outras páginas terminou sempre no regresso à *homepage*, antes de passar-se a uma data seguinte. Para referenciar uma versão na exposição virtual utilizou-se o URL da versão preservada.



ARQUIVO.PT

iha.fcsh.unl.pt/

2008

Fevereiro

19

03:56

Março 19 Fevereiro, 2008

Outubro

2009

2010

2012

2013

arquivo.pt/wayback/20080219035604/http://iha.fcsh.unl.pt/

HH INSTITUTO DE HISTÓRIA DA ARTE

Descobrir e Pensar a Cultura

INTRODUÇÃO | PUBLICAÇÕES | CURSOS | NEWSLETTER | CONTACTOS | OUTROS LINKS | + INFO | MAPA DO SITE | UNIDADE DE INVESTIGAÇÃO

APRESENTAÇÃO >>

OBJECTIVOS >>

ÓRGÃOS >>

MEMBROS >>

FUNCIONAMENTO >>

III CURSO MONOGRÁFICO

*A Porcelana Chinesa*

Dando continuidade a uma das linhas de trabalho do Instituto de História da Arte, iniciou-se no passado dia 8 de Novembro, na Casa Museu Dr. Anastácio Gonçalves, o III Curso Monográfico que este ano é dedicado à Porcelana da China. A leccionação do curso, organizado em vinte sessões, é assegurada pela Dra. Maria Antónia Pinto de Matos, Conservadora do Museu Nacional de Arte Antiga.

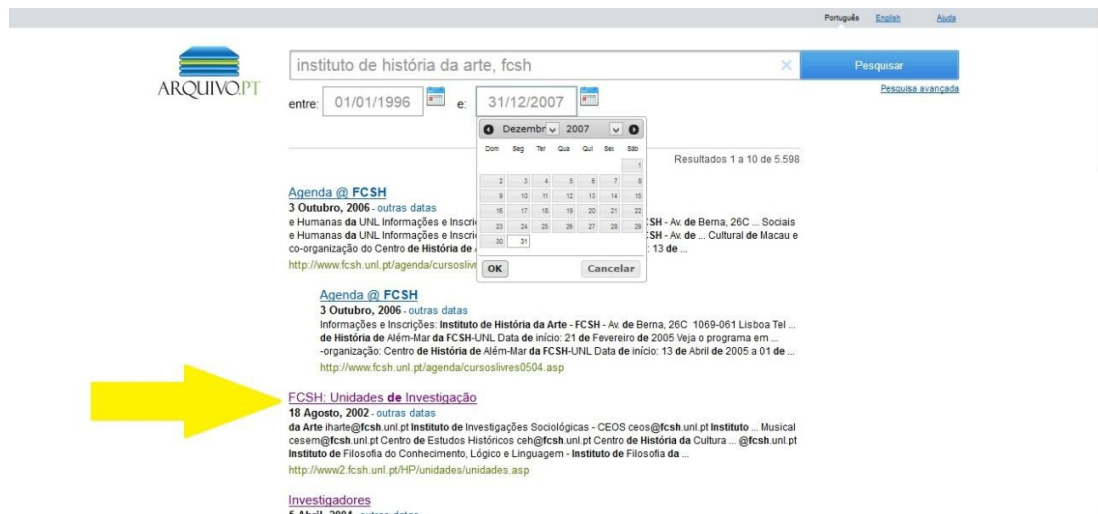
Fig. 2 – Interface do Arquivo.pt com apresentação cronológica (URL: <http://arquivo.pt/wayback/20080219035604/http://iha.fcsh.unl.pt/>).

A pesquisa procurou ser linear, de versão em versão, evitando seguir a pista de outras páginas do *site*. A título de exemplo, suponha-se que um sítio Web teve, a determinada altura da sua história, uma página que só durou dois anos, tendo sido retirada ao fim desse tempo; de pouco serve seguir essa pista para conhecer a história de todo o *site*. Por essa razão e para não perder a ligação entre as versões usou-se a homepage como referência em cada versão.

Neste processo de percorrer as versões dos sítios Web da FCSH, procurou-se informações para a história do *site*. Anotou-se características do *design* e das tecnologias presentes. Fez-se uma leitura dos conteúdos e de outras informações sobre a entidade representada no sítio Web. Finalmente, escolheu-se uma imagem para colocar na exposição *online*. Todos os apontamentos foram desta viagem no passado foram registados numa folha de recolha dedicada a cada um.

A função da imagem é ser um elemento visual de leitura imediata que recorda o passado da Faculdade ou de um centro de investigação. No caso de não haver imagens na primeira versão procurou-se nas versões seguintes. Não havendo imagens em condições nas versões seguintes, optou-se por produzir uma imagem, por escolher um pormenor, por melhorar a qualidade das imagens disponíveis.

A pesquisa por palavra ou por expressão foi utilizada como recurso complementar para descobrir novas pistas sobre a história dos *sites* e das entidades que os criaram. Segue-se um exemplo na Fig. 3.



Português English Ajuda

ARQUIVO.PT

instituto de história da arte, fcsh

Pesquisar

entre: 01/01/1996 e: 31/12/2007

Dezembro 2007

Dom Seg Ter Qua Qui Sex Sáb

1

2 3 4 5 6 7 8

9 10 11 12 13 14 15

16 17 18 19 20 21 22

23 24 25 26 27 28 29

30 31

OK Cancelar

Resultados 1 a 10 de 5.598

[Pesquisa avançada](#)

[Agenda @ FCSH](#)

3 Outubro, 2006 - outras datas

e Humanas da UNL. Informações e Inscri

e Humanas da UNL. Informações e Inscri

co-organização do Centro de História de

<http://www.fcsh.unl.pt/agenda/cursosivri>

[Agenda @ FCSH](#)

3 Outubro, 2006 - outras datas

Informações e Inscrições: Instituto de História da Arte - FCSH - Av. de Berna, 26C - 1069-061 Lisboa Tel ...

de História de Além-Mar da FCSH-UNL. Data de início: 21 de Fevereiro de 2005 Veja o programa em ...

-organização: Centro de História de Além-Mar da FCSH-UNL. Data de início: 13 de Abril de 2005 a 01 de ...

<http://www.fcsh.unl.pt/agenda/cursosivres0504.asp>

**FCSH: Unidades de Investigação**

18 Agosto, 2002 - outras datas

da Arte iharte@fcsh.unl.pt Instituto de Investigações Sociológicas - CEOS ceos@fcsh.unl.pt Instituto ... Musical

cesem@fcsh.unl.pt Centro de Estudos Históricos ce@fcsh.unl.pt Centro de História da Cultura ... @fcsh.unl.pt

Instituto de Filosofia do Conhecimento, Lógico e Linguagem - Instituto de Filosofia da ...

<http://www2.fcsh.unl.pt/HP/Unidades/unidades.asp>

[Investigadores](#)

5 Abril, 2004 - outras datas

Fig.3 – Pesquisa por expressão no Arquivo.pt, com especificação do limite de tempo (período anterior a 2008).

Assim, ao pesquisar “Instituto de História de Arte, FCSH” obtém-se uma lista de resultados com a informação de que, em 2002, esta unidade já existia e tinha uma página própria, inserida no sítio Web da Faculdade (Fig. 4). Desta página, existem 37 versões todas iguais entre 2000 e 2005. Não vem acrescentar muito sobre a história do sítio Web, mas é uma informação a ter em conta.



Fig. 4 – Página única do Instituto de História da Arte da FCSH, em 2002. (<http://arquivo.pt/wayback/20020614233300/http://www.fcsh.unl.pt/lha/default.htm>)

Continuando na lista de resultados, encontram-se ainda outros *sites* relacionados, como por exemplo, o do Núcleo de Estudos de Arte Contemporânea com o endereço [neac-ih.a.fcsh.unl.pt](http://neac-ih.a.fcsh.unl.pt).





Fig. 5 – Sítio Web do núcleo de Estudos de Arte Contemporânea do IHA. (<http://arquivo.pt/wayback/20050122003600/http://neac-iha.fcsh.unl.pt/>)

O processo de pesquisa e de análise dos sítios Web preservados no Arquivo.pt, pertencentes a entidades da FCSH, permitiu recolher informação sobre a presença online da Faculdade e de 23 das suas entidades, totalizando 24 itens, e passar à fase seguinte, ou seja, à montagem da exposição *online*.

### **3. Desenvolvimento de um sítio Web para a exposição *online***

#### **3.1 Escolha da plataforma para a exposição**

A escolha recaiu sobre o Wordpress. Em relação a este aspecto não houve discussão, porque desde o princípio o que se pretendia era um *software* amplamente utilizado, *open source*, com possibilidades de ser personalizado e desenvolvido. Por coincidência o Arquivo.pt estava a mudar o seu sítio Web para o Wordpress e, no âmbito do grupo Investiga XXI, foram criadas páginas para apresentar os projectos. Assim, foi possível aproveitar sinergias e ter formação sobre este sistema, beneficiando do apoio especializado dos informáticos da equipa.

Escolhido o sistema, optou-se por utilizar apenas a versão gratuita do serviço [Wordpress.com](http://Wordpress.com). Havia a possibilidade de criar um ambiente local de desenvolvimento com a versão profissional, mas isso foi adiado para trabalho futuro por várias razões. O pouco tempo disponível – menos de 3 meses – foi uma delas, mas não a principal. Pretendeu-se mostrar o que é possível fazer com uma versão básica que não exige conhecimentos técnicos especiais nem consome recursos incomportáveis para pequenas organizações.

Como já se disse no início desta apresentação, quis dar-se a este projecto um carácter exemplar, mostrando que está ao alcance de qualquer profissional da informação criar algo semelhante na sua própria organização.

#### **3.2 Criação de uma estrutura para a exposição**

O que pode dizer-se, desde logo, é que houve um processo de depuração dos conteúdos a inserir na exposição, tendo sido excluídas algumas ideias iniciais. Este discernimento teve o contributo importante do grupo de trabalho [Investiga XXI](#), através das suas apreciações.

O esquema final ficou da seguinte forma:

- Página inicial (entrada na exposição)
- Página do item

- Perguntas frequentes
- Sobre o projecto
- Contactos

Nesta fase definiu-se como unidade de referência o item. A exposição consiste num conjunto de itens, um por página. Em cada um expõe-se a memória *online* de uma entidade ou unidade organizacional, das que foram anteriormente pesquisadas e analisadas nas versões preservadas do Arquivo.pt (unidades I&D). No seu conjunto, expõem a presença *online* da FCSH entre 1996 e 2016.

De parte ficou a ideia inicial que consistia em criar duas colecções temáticas, uma página de recomendações e um blog. Desde que haja número de itens suficientes – e nesta fase não há – uma página de colecções é muito útil. As recomendações foram substituídas por uma página de “perguntas frequentes” e o *blog* foi dissociado do sítio Web da exposição. Em seu lugar foi utilizado para publicação de tutoriais e conteúdos de apoio o *blog* do autor (<http://preservacaodigital.wix.com/guia/blogue>).

e uma página no Facebook (<facebook.com/gpsiglas>).

### 3.3 Adequação da interface da exposição virtual

A interface baseou-se num *template* disponível para a versão gratuita. Neste caso foi escolhido um que tivesse uma página inicial baseada em imagens – o “Apostrophe 2”. Desde que o projecto foi imaginado, a intenção foi provocar a descoberta da memória a partir de uma imagem.

Deve dizer-se que a escolha do *template* adequado é uma tarefa morosa, pois nenhuma corresponde perfeitamente ao que se precisa. É sempre necessário um trabalho de adequação. As funcionalidades são facilmente personalizáveis na interface de administração do Wordpress. Existe documentação disponível para as tarefas a realizar, mesmo para quem não é especialista.

O *design* da página foi limpo o mais possível de todos os elementos secundários, que desviam a atenção do conteúdo que se pretende expor. Evitou-se botões desnecessários, simplificou-se a navegação, evitou-se abertura de novas janelas. Descartou-se a criação de elementos gráficos próprios (*icon*, *banner* no cabeçalho da página e outros elementos de *design*), uma vez que essa tarefa especializada não é o principal objectivo da exposição.

As imagens escolhidas, essas sim, foram tratadas num editor de imagem para poderem ser apresentadas na página de entrada e na página de cada item.



Os sítios Web antigos tem poucas imagens e as que aparecem geralmente têm pouca qualidade. Este facto foi uma dificuldade considerável para o objectivo de basear os pontos de entrada na exposição em imagens. Foram feitos cortes, escolhidos pormenores de páginas obtidas por *printscreens* e ainda utilizadas ainda imagens de outras fontes.

Quanto ao design, o resultado obtido para a página principal foi o seguinte (Fig. 6):

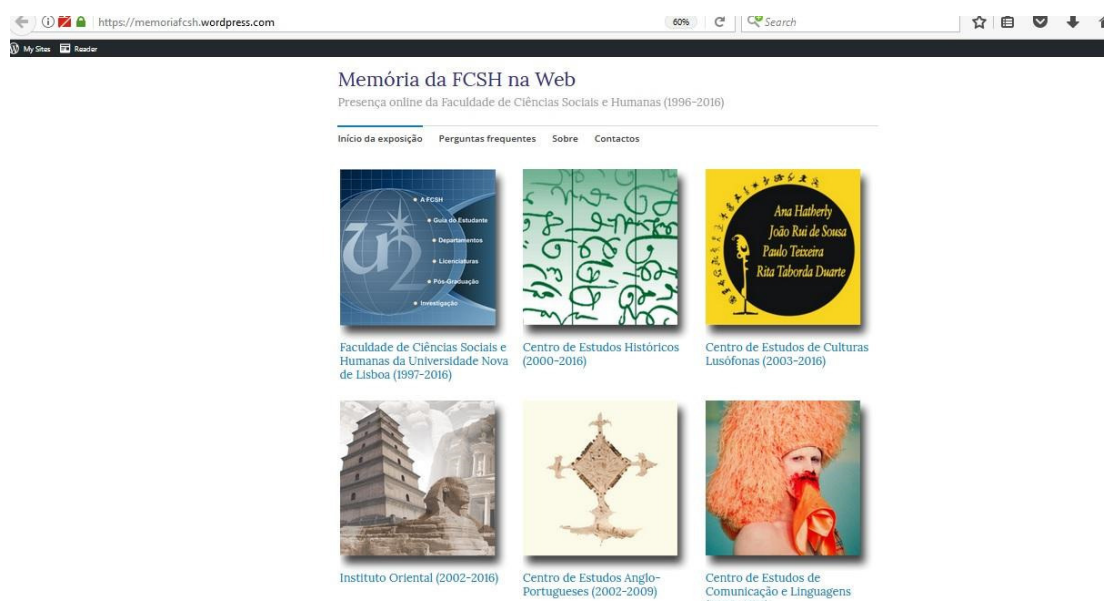


Fig. 6 – Página de entrada da exposição *online* (<https://memoriafcsch.wordpress.com>).

### 3.4 Apresentação dos conteúdos – memória dos sítios Web

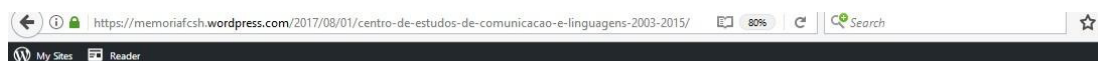
O requisito da Usabilidade esteve especialmente presente nesta fase de desenvolvimento do sítio Web. Os elementos a incluir numa página de apresentação acerca da memória de um sítio Web preservado deveriam ser reduzidos ao essencial, simples na estrutura e claríssimos na informação passada ao utilizador.

Valeu o apoio do orientador do projecto que é versado em questões da Usabilidade e escrita para a Web. Quer isto dizer, que a criação de um projecto semelhante, independentemente da dimensão, deve contar com o aconselhamento de alguém especializado na comunicação pela Web.

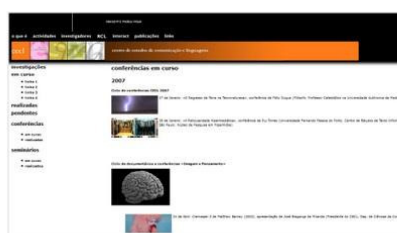
A página do item apresentou a seguinte estrutura, como se pode ver na imagem (Fig. 7):

- título:
- imagem de destaque

- sinopse: memória do sítio Web
- endereços
- momentos para recordar
- *sites* relacionados



## Centro de Estudos de Comunicação e Linguagens (2003-2016)



Pormenor da página "Actividades" ([cecl.unl.com.pt](http://cecl.unl.com.pt), 2003)

### Memória do sítio Web

A [primeira versão preservada é de 2003](#). Alguns elementos gráficos estão em Flash, o que limita a sua reprodução em vários dispositivos actuais. Em 2008, surge com um *design* ligeiramente diferente.

Em 2014, sítio Web, já com um *design* diferente, o sítio Web informa que o CECL [passou a integrar o centro inter-universitário CIC.DIGITAL](#), tornando-se seu pólo

Fig. 7 – Post dedicado ao Centro de Estudos de Comunicação e Linguagens (2003-2016) (<https://memoriafcsch.wordpress.com/2017/08/01/centro-de-estudos-de-comunicacao-e-linguagens-2003-2015/>).

Nesta página é possível ver a assumpção de algumas ideias expressas anteriormente. A organização ou entidade é o ponto de partida para a organização dos conteúdos relativos a um item. A entidade cria o sítio Web e utiliza-o como canal de comunicação. O sítio Web, independentemente do *hostname* ou do endereço, é um repositório de memórias. Aqui, está ilustrada a perspectiva organizacional de que se falou.

A questão dos diversos *hostnames* foi resolvida como se vê: o *site* Centro de Estudos de Comunicação e Linguagens teve apenas um endereço ao longo dos anos, mas o Instituto de História da Arte teve dois.

Os “Momentos para recordar” são simples *links* que lançam o visitante na interface do Arquivo.pt.

Quanto aos *sites* relacionados, é um aspecto para desenvolvimento futuro. Nesta página é utilizada a funcionalidade das categorias do Wordpress para gerar alguma dinâmica no fim da página.

Na perspectiva da escrita para a Web, o texto com a “Memória do sítio Web” foi reduzido a um limite de 150 palavras e alguns *links* que remetem o visitante da exposição para o Arquivo.pt. Por analogia com uma exposição física o texto pode comparar-se à placa que se coloca ao lado dos objectos com a sua descrição.

### 3.5 Disseminação do projecto

A comunicação do projecto em diversos contextos foi uma parte importante no seu desenvolvimento. Serviram para divulgar o projecto, mas também para defini-lo melhor.

As apresentações tiveram a forma e o contexto seguinte:

- Encontro Nacional da FCCN, realizado em 20 de Abril de 2017 na Universidade de Trás-os-Montes Alto Douro.
- Aula aberta na Faculdade de Ciências Sociais e Humanas, inserido numa formação sobre o Arquivo.pt para alunos de comunicação social, a 10 de Maio de 2017.
- Vídeo de apresentação gravado no estúdio da FCCN para divulgação no sítio Web do Arquivo.pt e nas redes sociais.
- Apresentação do projecto aos representantes das unidades de investigação da Faculdade, numa reunião de trabalho.
- E ainda, tutoriais no *blog* do autor sobre como pesquisar sobre a memória da Faculdade no Arquivo.pt, estratégia para desenvolvimento futuro ([@gpsiglas](#)).

## 4. Trabalho futuro

### 4.1 Descrição e curadoria dos sítios Web da Faculdade

Durante o desenvolvimento deste projecto investigou-se sobre a forma de fazer um ficha para descrever um sítio Web. Esta acabou por não ser incluída na exposição.

Subjacente a esta ideia está perspectiva do bibliotecário, habituado a que tudo tenha uma catalogação, como acontece com outros recursos, tais como livros, revistas, documentos electrónicos de vários tipos e em vários formatos.

Há bibliotecas que inserem registos de sítios Web nos seus catálogos, como por exemplo, a da Universidade de Yale, que utiliza o conceito de "integrating resource" para diferenciar um sítio Web de uma publicação periódica (<http://web.library.yale.edu/cataloging/e-resources/updating-websites>). No entanto, fazem-no em relação aos sítios Web correntes.

Para criar um catálogo de sítios Web do passado, limitado a uma instituição como a Faculdade, teria de ser utilizado um modelo semelhante ao do projecto [Archivit-it](#), nascido a partir do Archive.org. Este supõe a criação de um serviço, pelo que só interessa como perspectiva.

Está ao alcance do gestores de informação pelo menos pensar no assunto e procurar modelos de descrição de sítios Web que incluam as versões preservadas, que permitam cruzar dados e caracterizar a presença de uma instituição na Web.

Foi o que se pretendeu ao considerar os metadados necessários para descrição estruturada de um sítio Web (Tab. 2):

Tabela 2: Correspondência entre os elementos Dublin Core e os metadados para a ficha do *site* no projecto da exposição *online*

FICHA DO SÍTIO WEB	ELEMENTOS DC	LABEL OU ETIQUETA	DADOS A RECOLHER
Dados relativos ao conteúdo	DC.Title	Nome	O nome dado ao <i>site</i>
	DC.Subject	Categorias	Assuntos representados por <i>keywords</i>
	DC.Subject	Tags	<i>idem</i>
	DC.Description	Descrição	Contém xyz. É composto por. Existem X versões no Arquivo.pt. Teve vários designs. Utiliza tecnologias xyz, sistemas xyz. Tem funcionalidades xyz. O que não está a funcionar xyz.
	DC.Relation	<i>Sites</i> relacionados no Arquivo.pt	<i>Site</i> com algum tipo de relação, como por exemplo, <i>sites</i> de projecto, eventos, redes de cooperação.

	DC.Coverage	Cobertura temporal	Período temporal delimitado que se quer descrever na exposição <i>online</i>
Dados relativos à propriedade de	DC.Creator	Autor	Autor: a entidade principal responsável pela criação do <i>site</i> .
	DC.Contributor	Co-autor	Co-autor: a entidade secundária que prestou algum apoio (patrocinador, por ex.).
	DC.Publisher	Editor	Editor: a entidade principal responsável pela comunicação através do <i>site</i> .
	DC.rights	Direitos	<i>Disclaimer</i> . Texto ou url da página no Arquivo.pt
Versão de referência	DC.Date	Data (1ª versão)	Data da 1ª versão de um <i>site</i> conservada no Arquivo.pt
	DC.Identifier	URI-Arquivo.pt	URL que aponta para a 1ª versão no Arquivo.pt

Referência: <http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=elements#H2>

#### 4.2 Integração do projecto nas redes de comunicação da FCSH

Desde o início, foi dado conhecimento à Faculdade que o projecto estava a realizar-se com o Arquivo.pt e foi pedido o apoio institucional, ou seja, um reconhecimento da utilidade do projecto e facilidades no acesso a informação junto dos centros. Esse apoio foi dado por intermédio Biblioteca e ainda pelos responsáveis das áreas da Investigação e da Comunicação. Da parte dos serviços de informática também houve um bom acolhimento e apoio.

Reconhece-se a oportunidade do projecto para a comemoração dos 40 anos da Faculdade, em Novembro de 2017.

A forma e os modos de divulgação no espaço físico da Faculdade e nos canais de comunicação será tratada no arranque do novo ano lectivo.

#### 4.3 Desenvolvimento da exposição *online*

A transposição do projecto para um contexto institucional como o da Faculdade supõe um aumento do número de itens apresentados na exposição, para que todas as entidades que tiveram os seus sítios Web no passado se sintam representadas na

exposição. Neste protótipo de exposição estão incluídos 23 entidades e a Faculdade. No entanto, existem mais.

A longo prazo seria desejável incluir todos os *sites* de projectos e eventos relacionados, incluindo aqueles que foram sugeridos para preservação e que, certamente, dentro de um ano, estarão no Arquivo.pt. Há ainda os *sites* dos Departamentos que foram criados dentro do site da Faculdade e os outros de grupos de alunos (Associação de Estudantes, tunas e grupo de teatro, etc).

A comunicação com os representantes de todas estas entidades é importante para melhorar o conteúdo e para divulgar a exposição. Aí está um dos aspectos do desenvolvimento do trabalho. Não se sabe qual será a resposta. É expectável obter contributos (“testemunhos”) para serem publicados no *blog*.

## **5. Observações finais e aspectos a reter para um gestor da informação**

O *Web archiving* é uma actividade específica que ainda não é habitual entre as tarefas de curadoria da informação numa instituição. Este projecto foi uma oportunidade muito interessante de tomar contacto com essa realidade e compreender os seus desafios.

O modelo de orientação foi excelente, porque proporcionou um desenvolvimento de trabalho em grupo e ainda um acompanhamento individual extraordinário, muito enriquecedor do ponto de vista formativo.

O tempo que durou o projecto foi pouco para desenvolver alguns aspectos fundamentais, nomeadamente uma formação ainda mais aprofundada sobre o Arquivo.pt.

As bibliotecas universitárias são potencialmente a melhor mediação para levar o Arquivo.pt às Universidades. Imagine-se o que seria poder ter um serviço semelhante ao Archive-it, em que as Universidades criam as suas próprias colecções de sítios Web, a partir do Arquivo.pt.

O resultado do projecto demonstra que é possível promover a preservação dos sítios Web valendo-se das competências normais de um gestor de informação.